

MAXIMUM LIKELIHOOD DEGREE OF FERMAT HYPERSURFACES VIA EULER CHARACTERISTICS

BOTONG WANG

ABSTRACT. Maximum likelihood degree of a projective variety is the number of critical points of a general likelihood function. In this note, we compute the Maximum likelihood degree of Fermat hypersurfaces. We give a formula of the Maximum likelihood degree in terms of the constants $\beta_{\mu,\nu}$, which is defined to be the number of complex solutions to the system of equations $z_1^\nu = z_2^\nu = \cdots = z_\mu^\nu = 1$ and $z_1 + \cdots + z_\mu + 1 = 0$.

1. INTRODUCTION

The maximum likelihood estimate is a fundamental problem in statistics. Maximum likelihood degree is the number of potential solutions to the maximum likelihood estimation problem on a projective variety. When the variety is smooth, Huh [H] showed that the Maximum likelihood degree is indeed a topological invariant. If the variety is a general complete intersection, the maximum likelihood degree is computed in [CHKS] (see also [HS]).

In a recent preprint [AAGL], Agostini, Alberelli, Grande and Lella studied the maximum likelihood degree of Fermat hypersurfaces. They obtained formulas for the maximum likelihood degree of a few special families of Fermat surfaces. However, their approach is through a case-by-case study.

In this note, we propose to compute the Maximum likelihood degree of Fermat hypersurfaces in a more systematic way via topological method. In general, the formula given in [CHKS] does not work for all the Fermat hypersurfaces, because the intersection of hypersurfaces

$$\{x_0^d + x_1^d + \cdots + x_n^d = 0\} \cap \{x_0 + x_1 + \cdots + x_n = 0\} \subset \mathbb{P}^n$$

may not be transverse. We will compute the error terms introduced by the non-transverse intersections. The main ingredient is Milnor's result on the topology of isolated hypersurfaces singularities. This topological approach is closely related to the approach of [BW] and [RW]. In fact, for an isolated hypersurface singularity, the Euler obstruction is up to a sign equal to the Milnor number plus one. So we essentially apply the ideas of [BW] and [RW] to these particular examples.

First, let us recall the definition of Maximum likelihood degree. Let \mathbb{P}^n be the n -dimensional complex projective space with homogeneous coordinates (x_0, x_1, \dots, x_n) . Denote the coordinate plane $\{x_i = 0\} \subset \mathbb{P}^n$ by H_i , and the hyperplane $\{x_0 + x_1 + \cdots + x_n = 0\}$ by H_+ . Let the index set $\Lambda = \{0, 1, \dots, n, +\}$, and let $\mathcal{H} = \bigcup_{\lambda \in \Lambda} H_\lambda$. Let $X \subset \mathbb{P}^n$ be a complex projective variety. Denote the smooth locus of X by X_{reg} . The **Maximum likelihood degree** of X is defined to be the number of critical points of

the likelihood function

$$l_u = \frac{x_0^{u_0} x_1^{u_1} \cdots x_n^{u_n}}{(x_0 + x_1 + \cdots + x_n)^{u_0 + u_1 + \cdots + u_n}}$$

on $X_{\text{reg}} \setminus \mathcal{H}$ for generic $(u_i)_{0 \leq i \leq n} \in \mathbb{Z}^{n+1}$.

Theorem 1.1. *Denote the Fermat hypersurface $\{x_0^d + x_1^d + \cdots + x_n^d = 0\} \subset \mathbb{P}^n$ by $F_{n,d}$, and denote its maximum likelihood degree by $\text{MLdeg}(F_{n,d})$. Then,*

$$(1) \quad \text{MLdeg}(F_{n,d}) = d + d^2 + \cdots + d^n - \sum_{0 \leq j \leq n-1} \binom{n+1}{j} \beta_{n-j,d-1}$$

where $\beta_{\mu,\nu}$ is the number of complex solutions of the system of equations

$$\begin{aligned} z_1^\nu &= z_2^\nu = \cdots = z_\mu^\nu = 1 \\ z_1 + \cdots + z_\mu + 1 &= 0. \end{aligned}$$

When μ or ν is small, $\beta_{\mu,\nu}$ can be easily calculated. For example,

$$(2) \quad \beta_{\mu,1} = 0.$$

$$(3) \quad \beta_{1,\nu} = \begin{cases} 0 & \text{if } \nu \text{ is odd,} \\ 1 & \text{if } \nu \text{ is even.} \end{cases}$$

$$(4) \quad \beta_{2,\nu} = \begin{cases} 2 & \text{if } \nu \text{ is divisible by 3,} \\ 0 & \text{otherwise.} \end{cases}$$

With these calculations, we recover all the closed formulas in [\[AAGL\]](#).

Corollary 1.2.

$$(5) \quad \text{MLdeg}(F_{n,2}) = 2^{n+1} - 2$$

$$(6) \quad \text{MLdeg}(F_{2,d}) = \begin{cases} d^2 + d & \text{if } d \equiv 0, 2 \pmod{6}, \\ d^2 + d - 3 & \text{if } d \equiv 3, 5 \pmod{6}, \\ d^2 + d - 2 & \text{if } d \equiv 4 \pmod{6}, \\ d^2 + d - 5 & \text{if } d \equiv 1 \pmod{6}. \end{cases}$$

When ν is a power of a prime number, we have formulas to compute $\beta_{\mu,\nu}$. Equivalently, when $d-1$ is a power of a prime number, we have closed formulas for $\text{MLdeg}(F_{n,d})$. In fact, by a straight forward computation one can deduce the following corollary from Theorem 1.1 and Proposition 4.2.

Corollary 1.3. *Suppose $d-1 = p^r$, where p is a prime number and r is a positive integer. Then*

$$\text{MLdeg}(F_{n,d}) = d + d^2 + \cdots + d^n - \frac{1}{d-1} \sum \frac{(n+1)!}{(n+1-p(s_1+\cdots+s_k))! \cdot ((s_1)! \cdots (s_k)!)^p}$$

where $k = \frac{d-1}{p}$ and the sum is over all nonnegative integers s_1, \dots, s_k with $1 \leq s_1 + \dots + s_k \leq \frac{n+1}{p}$.

To find a general formula for $\beta_{\mu,\nu}$ would be a very hard question in number theory and combinatorics. In fact, determining when $\beta_{\mu,\nu} \neq 0$ had been an open question for a long time, and it was solved by Lam and Leung [LL] in 2000.

Since the Fermat hypersurface $F_{n,d}$ is smooth, by [H] $\text{MLdeg}(F_{n,d})$ is equal to the signed Euler characteristic $\chi(F_{n,d} \setminus \mathcal{H})$. In section 2, we will compute $\chi(F_{n,d} \setminus \mathcal{H})$, and we will postpone the technical calculation of the Milnor numbers to section 3. In the last section, we will briefly discuss what we know about the constants $\beta_{n,d}$.

Acknowledgement. We thank Jiu-Kang Yu and Zhengpeng Wu for helpful discussions about the constants $\beta_{\mu,\nu}$.

2. COMPUTING THE EULER CHARACTERISTICS

By the following theorem of Huh [H], we reduce the problem of computing $\text{MLdeg}(F_{n,d})$ to computing $\chi(F_{n,d} \setminus \mathcal{H})$. Recall that in \mathbb{P}^n , $\mathcal{H} = \bigcup_{\lambda \in \Lambda} H_\lambda$ is the union of all coordinate hyperplanes and the hyperplane $H_+ = \{x_0 + x_1 + \dots + x_n = 0\}$.

Theorem 2.1 (Huh, [H]). *If $X \subset \mathbb{P}^n$ is a subvariety such that $X \setminus \mathcal{H}$ is smooth, then*

$$\text{MLdeg}(X) = (-1)^{\dim(X)} \chi(X \setminus \mathcal{H}).$$

Since the Euler characteristic is additive for algebraic varieties, by the inclusion-exclusion principle,

$$(7) \quad \chi(X \setminus \mathcal{H}) = \sum_{0 \leq i \leq n} \sum_{\substack{\Lambda' \subset \Lambda \\ |\Lambda'|=i}} (-1)^i \chi(X \cap H_{\Lambda'})$$

where $H_{\Lambda'} = \bigcap_{\lambda \in \Lambda'} H_\lambda$.

The Fermat hypersurface $F_{n,d} = \{x_0^d + x_1^d + \dots + x_n^d = 0\}$ is invariant under any permutation of the coordinates. Therefore, (7) can be written as

$$(8) \quad \chi(F_{n,d} \setminus \mathcal{H}) = \sum_{0 \leq i \leq n} (-1)^i \left(\binom{n+1}{i} \chi(F_{n,d} \cap V^i) + \binom{n+1}{i-1} \chi(F_{n,d} \cap W^i) \right)$$

where $V^i = \bigcap_{0 \leq j \leq i-1} H_j$ and $W^i = H_+ \cap \bigcap_{0 \leq j \leq i-2} H_j$ ($W^0 = \emptyset$ and $W^1 = H_+$).

$F_{n,d} \cap V^i$ is a smooth hypersurface in \mathbb{P}^{n-i} of degree d . Euler characteristics of such hypersurfaces only depend on $n-i$ and d , and they are calculated in [D] Chapter 5, (3.7). However, it turns out that we don't have to compute each of these Euler characteristics. For now, we simply denote the Euler characteristic of a smooth degree d hypersurfaces in \mathbb{P}^m by $e_{m,d}$. In particular,

$$(9) \quad \chi(F_{n,d} \cap V^i) = e_{n-i,d}.$$

$F_{n,d} \cap W^i$ is a possibly singular hypersurface in W^i for $1 \leq i \leq n$. In fact, $F_{n,d} \cap W^i$ is isomorphic to the intersection of the Fermat hypersurface $F_{n-i+1,d} \subset \mathbb{P}^{n-i+1}$ and the hyperplane $\{x_0 + x_1 + \dots + x_{n-i+1} = 0\}$. Using Lagrange multiplier method, one can easily see that all the singular points of $F_{n,d} \cap W^i$ are isolated and there are exactly $\beta_{n-i+1,d-1}$

many of them. The Euler characteristics of such hypersurfaces can be computed using Milnor numbers.

Theorem 2.2. [D, Chapter 5 (4.4)] *For any singular point P of $F_{n,d} \cap W^i$ we can define the Milnor number $\mu(F_{n,d} \cap W^i, P)$ by considering $F_{n,d} \cap W^i$ as a hypersurface of W^i . Then,*

$$(10) \quad \chi(F_{n,d} \cap W^i) = e_{n-i,d} + (-1)^{n-i} \sum_P \mu(F_{n,d} \cap W^i, P)$$

where the sum is over all the singular points P of $F_{n,d} \cap W^i$.

Proposition 2.3. *For any singular point P of $F_{n,d} \cap W^i$,*

$$(11) \quad \mu(F_{n,d} \cap W^i, P) = 1.$$

We will postpone the proof of the proposition to next section. The next corollary follows immediately from (10) and (11).

Corollary 2.4.

$$(12) \quad \chi(F_{n,d} \cap W^i) = e_{n-i,d} + (-1)^{n-i} \beta_{n-i+1,d-1}.$$

Now, combining (8), (9) and (12), we have

$$(13) \quad \chi(F_{n,d} \setminus \mathcal{H}) = \sum_{0 \leq i \leq n} (-1)^i \left(\binom{n+1}{i} e_{n-i,d} + \binom{n+1}{i-1} (e_{n-i,d} + (-1)^{n-i} \beta_{n-i+1,d-1}) \right)$$

Since $\binom{n+1}{i} + \binom{n+1}{i-1} = \binom{n+2}{i}$, (13) is equivalent to

$$(14) \quad \chi(F_{n,d} \setminus \mathcal{H}) = \sum_{0 \leq i \leq n} (-1)^i \binom{n+2}{i} e_{n-i,d} + \sum_{1 \leq i \leq n} (-1)^n \binom{n+1}{i-1} \beta_{n-i+1,d-1}.$$

Suppose X is a general hypersurface of degree d in \mathbb{P}^n . Then (7) implies that

$$(15) \quad \chi(X \setminus \mathcal{H}) = \sum_{0 \leq i \leq n} (-1)^i \binom{n+2}{i} e_{n-i,d}$$

The maximum likelihood degree of a general hypersurfaces is well-understood.

Proposition 2.5. [HS, 1.11] *The maximum likelihood of a general degree d hypersurfaces in \mathbb{P}^n is equal to $d + d^2 + \cdots + d^n$.*

Combining the proposition, (15) and Theorem 2.1, we have

$$(16) \quad \begin{aligned} \sum_{0 \leq i \leq n} (-1)^i \binom{n+2}{i} e_{n-i,d} &= \chi(X \setminus \mathcal{H}) \\ &= (-1)^{n-1} \text{MLdeg}(X) \\ &= (-1)^{n-1} (d + d^2 + \cdots + d^n) \end{aligned}$$

Therefore, (14) is equivalent to

$$(17) \quad \chi(F_{n,d} \setminus \mathcal{H}) = (-1)^{n-1} (d + d^2 + \cdots + d^n) + \sum_{1 \leq i \leq n} (-1)^n \binom{n+1}{i-1} \beta_{n-i+1,d-1}$$

Again, by Theorem 2.1, we have

$$(18) \quad \text{MLdeg}(F_{n,d}) = d + d^2 + \cdots + d^n - \sum_{1 \leq i \leq n} \binom{n+1}{i-1} \beta_{n-i+1,d-1}$$

which is the statement of Theorem 1.1.

3. THE MILNOR NUMBERS

We prove Proposition 2.3 in this section.

For the geometric meaning of Milnor number, we refer to [D, Chapter 3]. Here we compute the Milnor numbers using Jacobian ideals. Denote the ring of germs of holomorphic functions at $0 \in \mathbb{C}^l$ by \mathcal{O} . Let $f \in \mathcal{O}$ be a nonzero germ of holomorphic function such that the germ of hypersurface $f^{-1}(0)$ has an isolated singularity at the origin $0 \in \mathbb{C}^l$. The Jacobian ideal of f , denoted by J_f is defined by

$$J_f = \left(\frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_l} \right) \subset \mathcal{O}$$

where z_1, \dots, z_l are the coordinates of \mathbb{C}^n .

Theorem 3.1. [D, Chapter 3, (2.7)] *The Milnor number of $f^{-1}(0)$ at the origin, denoted by $\mu(f^{-1}(0), 0)$, is given by the formula*

$$(19) \quad \mu(f^{-1}(0), 0) = \dim_{\mathbb{C}} \mathcal{O} / J_f.$$

Recall that $W^i = \{x_0 = x_1 = \cdots = x_{i-2} = x_0 + x_1 + \cdots + x_n = 0\} \subset \mathbb{P}^n$. Denote $y_j = x_{i-1+j}$, $0 \leq j \leq n-i+1$. Then the intersection $F_{n,d} \cap W^i$ is isomorphic to the intersection

$$\{y_0^d + y_1^d + \cdots + y_{n-i+1}^d = 0\} \cap \{y_0 + y_1 + \cdots + y_{n-i+1} = 0\}$$

in \mathbb{P}^{n-i+1} . Without loss of generality, we can work on the affine space $y_0 \neq 0$, and rewrite the intersection in affine coordinates

$$\{1 + \bar{y}_1^d + \cdots + \bar{y}_{n-i+1}^d = 0\} \cap \{1 + \bar{y}_1 + \cdots + \bar{y}_{n-i+1} = 0\}.$$

Here we use \bar{y}_j to denote the corresponding affine coordinate of y_j , that is, $\bar{y}_j = y_j/y_0$. Suppose $(\xi_1, \dots, \xi_{n-i+1})$ is a singular point of the above intersection. Then by Lagrange multiplier method,

$$(20) \quad \xi_1^{d-1} = \xi_2^{d-1} = \cdots = \xi_{n-i+1}^{d-1} = 1.$$

We can eliminate \bar{y}_{n-i+1} by $\bar{y}_{n-i+1} = 1 - \bar{y}_1 - \cdots - \bar{y}_{n-i}$. On this affine chart, $F_{n,d} \cap W^i$ is isomorphic to the hypersurface $\{f = 0\}$ in \mathbb{C}^{n-i} , where

$$(21) \quad f = 1 + \bar{y}_1^d + \cdots + \bar{y}_{n-i}^d + (1 - \bar{y}_1 - \cdots - \bar{y}_{n-i})^d.$$

Let $z_j = \bar{y}_j - \xi_j$. Then

$$(22) \quad f = 1 + (z_1 + \xi_1)^d + \cdots + (z_{n-i} + \xi_{n-i})^d + (\xi_{n-i+1} - z_1 - \cdots - z_{n-i})^d.$$

Proposition 3.2. *In the local ring \mathcal{O} , the Jacobian ideal $J_f = (\frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_{n-i}})$ is equal to the maximal ideal $(z_1, z_2, \dots, z_{n-i})$.*

Proof. Notice that $\xi_j^{d-1} = 1$ for all $1 \leq j \leq n - i + 1$. Therefore,

$$\begin{aligned} \frac{\partial f}{\partial z_j} &= \frac{d(d-1)}{2} \cdot \xi_j^{d-2} z_j + \frac{d(d-1)}{2} \cdot \xi_j^{d-2} (z_1 + \cdots + z_{n-i}) + \text{higher degree terms} \\ &= \frac{d(d-1)}{2} \cdot \xi_j^{d-2} (z_1 + \cdots + z_{j-1} + 2z_j + z_{j+1} + \cdots + z_{n-i}) + \text{higher degree terms} \end{aligned}$$

By Nakayama's lemma, we only need to show that the vectors $z_1 + \cdots + z_{j-1} + 2z_j + z_{j+1} + \cdots + z_{n-i}$, $1 \leq j \leq n - i$ span the whole vector space $\mathbb{C}z_1 \oplus \mathbb{C}z_2 \oplus \cdots \oplus \mathbb{C}z_{n-j}$. By adding all such vectors together, we see $z_1 + z_2 + \cdots + z_{n-i}$ is contained in their span. Thus

$$z_j = (z_1 + \cdots + z_{j-1} + 2z_j + z_{j+1} + \cdots + z_{n-i}) - (z_1 + z_2 + \cdots + z_{n-i})$$

is in the span. \square

Now, Proposition 2.3 follows from Theorem 3.1 and Proposition 3.2.

4. THE CONSTANTS $\beta_{\mu,\nu}$

Instead of working with the constants $\beta_{\mu,\nu}$, we define $\alpha_{\mu,\nu}$ to be the number of complex solutions to the system of equations

$$(23) \quad \begin{cases} z_1^\nu = z_2^\nu = \cdots = z_\mu^\nu = 1 \\ z_1 + z_2 + \cdots + z_\mu = 0. \end{cases}$$

Then clearly $\beta_{\mu,\nu} = \frac{1}{\nu} \cdot \alpha_{\mu+1,\nu}$. The advantage of working with $\alpha_{\mu,\nu}$ is that their defining equations have better symmetry.

We would like to answer the following question.

Question. Give a formula for $\alpha_{\mu,\nu}$ in terms of μ and the prime factorization of ν .

This is definitely a very hard question. The work of Lam and Leung gives a necessary and sufficient condition of $\alpha_{\mu,\nu} \neq 0$.

Theorem 4.1. [LL] *Suppose $\nu = p_1^{a_1} \cdots p_l^{a_l}$ is the prime factorization. Then $\alpha_{\mu,\nu} \neq 0$ if and only if $\mu \in \mathbb{Z}_{\geq 0} \cdot p_1 + \cdots + \mathbb{Z}_{\geq 0} \cdot p_l$.*

When $\nu = p^r$ has only one prime factor, we can give a formula of $\alpha_{\mu,\nu}$. In this case, suppose (z_1, \dots, z_μ) is a solution to (23). Then the collection $\{z_1, \dots, z_\mu\}$ can be divided into groups of p elements such that each group is a rotation of $1, e^{2\pi i/p}, \dots, e^{2(p-1)\pi i/p}$. Therefore, if p does not divide μ , then $\alpha_{\mu,\nu} = 0$. If p divides μ , then

$$(24) \quad \alpha_{\mu,\nu} = \sum \frac{\mu!}{((s_1)!(s_2)! \cdots (s_k)!)^p}$$

where $k = \mu/p$, and the sum is over all $s_1, \dots, s_k \in \mathbb{Z}_{\geq 0}$ such that $s_1 + \cdots + s_k = \mu/p$. Since $\beta_{\mu,\nu} = \frac{1}{\nu} \cdot \alpha_{\mu+1,\nu}$, we can translate (24) into a statement about $\beta_{\mu,\nu}$.

Proposition 4.2. *Suppose $\nu = p^r$, where p is a prime number and r is a positive integer. Then $\beta_{\mu,\nu} = 0$ when p does not divide $\mu + 1$, and when p divides $\mu + 1$*

$$(25) \quad \beta_{\mu,\nu} = \frac{1}{\nu} \sum \frac{(\mu + 1)!}{((s_1)!(s_2)! \cdots (s_k)!)^p}$$

where $k = \nu/p$, and the sum is over all $s_1, \dots, s_k \in \mathbb{Z}_{\geq 0}$ such that $s_1 + \dots + s_k = \frac{\mu+1}{p}$.

Suppose $\nu = p^r q^s$ has two distinct prime factors, and suppose (z_1, \dots, z_μ) is a solution to (23). Then by [LL, Corollary 3.4], the collection $\{z_1, \dots, z_\mu\}$ can be divided into groups of p or q elements such that each group is a rotation of $1, e^{2\pi i/p}, \dots, e^{2(p-1)\pi i/p}$ or a rotation of $1, e^{2\pi i/q}, \dots, e^{2(q-1)\pi i/q}$ respectively. However, this decomposition is not unique, and this is the main difficulty to find a formula for $\alpha_{\mu,\nu}$ in this case. Now, this is already a problem beyond our capability.

When ν has at least three distinct prime factors, the statement of [LL, Corollary 3.4] is not true any more. Therefore, the question becomes much harder and deeper.

REFERENCES

- [AAGL] D. Agostini, D. Alberelli, F. Grande and P. Lella, The maximum likelihood degree of Fermat hypersurfaces. arXiv:1404.5745. To appear in J. Algebr. Stat. [1](#), [2](#)
- [BW] N. Budur and B. Wang, Bounding the maximum likelihood degree. arXiv:1411.3486. To appear in Math. Res. Lett. [1](#)
- [CHKS] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels, The maximum likelihood degree. Amer. J. Math. 128 (2006), no. 3, 671–697. [1](#)
- [D] A. Dimca, Singularities and topology of hypersurfaces. Universitext. Springer-Verlag, New York, 1992. xvi+263 pp. ISBN: 0-387-97709-0 [3](#), [4](#), [5](#)
- [H] J. Huh, The maximum likelihood degree of a very affine variety. Compos. Math. 149 (2013), no. 8, 1245-1266. [1](#), [3](#)
- [HS] J. Huh and B. Sturmfels, Likelihood Geometry. Combinatorial algebraic geometry, 63-117, Lecture Notes in Math., 2108, Springer, Cham, 2014. [1](#), [4](#)
- [LL] T. Y. Lam and K. H. Leung, On vanishing sums of roots of unity. Journal of Algebra 224 (2000), 91-109 [3](#), [6](#), [7](#)
- [RW] J. Rodriguez and B. Wang, The maximum likelihood degree of rank 2 matrices via Euler characteristics. arXiv:1505.06536 [1](#)

E-mail address: `botong.wang@wis.kuleuven.be`

KU LEUVEN, DEPARTMENT OF MATHEMATICS, CELESTIJNENLAAN 200B, B-3001 LEUVEN, BELGIUM